MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS 1963 A
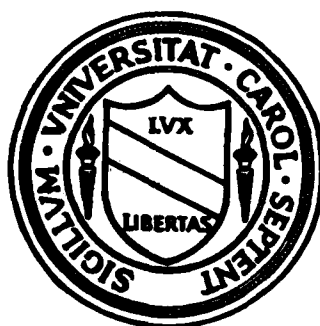
CENTER FOR STOCHASTIC PROCESSES

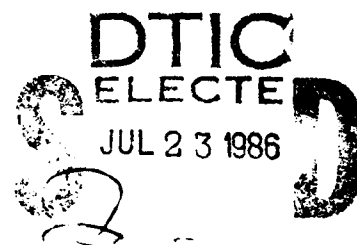Department of Statistics
University of North Carolina
Chapel Hill, North Carolina

NONPARAMETRIC SEQUENTIAL ESTIMATION OF ZEROS AND EXTREMA

OF REGRESSION FUNCTIONS

Wolfgang Härdle

and

Rainer Nixdorf

DTIC
ELECTE
JUL 2 3 1986

Technical Report No. 133

January 1986

8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620 85 C 0144			
8c. ADDRESS (City, State and ZIP Code) Bldg. 410 Bolling AFB, DC		10. SOURCE OF FUNDING NOS.			
		PROGRAM ELEMENT NO. 6.1102F	PROJECT NO. 2304	TASK NO. A5	WORK UNIT NO.

11. TITLE (Include Security Classification)
Nonparametric sequential estimation of zeros and extrema of regression functions

12. PERSONAL AUTHOR(S)
W. Hardle and R. Nixdorf

13a. TYPE OF REPORT technical	13b. TIME COVERED FROM 9/85 TO 8/86	14. DATE OF REPORT (Yr., Mo., Day) January 1986	15. PAGE COUNT 15

16. SUPPLEMENTARY NOTATION

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	Keywords: stochastic approximation, kernel regression, nonparametric regression, estimation of zeros and extremas.
XXXXX	XXXXXXXXXXX	XXXX	

NONPARAMETRIC SEQUENTIAL ESTIMATION OF ZEROS AND EXTREMA

OF REGRESSION FUNCTIONS

Wolfgang Härdle[*]
Johann Wolfgang Goethe - Universität
Fachbereich Mathematik
D-6000 Frankfurt/Main
and
Center for Stochastic Processes
University of North Carolina
Chapel Hill, NC

Rainer Nixdorf
Universität Stuttgart
Pfaffenwaldring 57
D-7000 Stuttgart 80

Abstract Let (X,Y), (X_1,Y_1), (X_2,Y_2),... be independent, identically distributed, bivariate random variables and let $m(x)=E(Y|X=x)$ be the regression curve of Y on X. In this paper we consider the estimation of zeros and extrema of the regression curve via stochastic approximation methods. We present consistency results of some sequential procedures and define termination rules providing fixed width confidence intervals for the parameters to be estimated.

AMS 1980 Subject classifications: Primary: 62L20

Secondary: 62J02, 62G05

Keywords and Phrases: Stochastic approximation, kernel regression, nonparametric regression, estimation of zeros and extremas.

1. Introduction

Let $(X,Y), (X_1, Y_1), (X_2, Y_2), \ldots$ be a sequence of independent, identically distributed, bivariate random variables with joint probability density function $f(x,y)$. In this paper we consider the sequential estimation of zeros and extrema of $m(x) = E(Y|X = x)$ using a combination of the nonparametric kernel and stochastic approximation methods. The structure of our sampling scheme is different from the one considered by Robbins and Monro (1951) since the experimenter, observing the bivariate data, has no control over the design variables $\{X_i\}$, as is assumed in classical stochastic approximation algorithms.

The proposed sequential procedure is based on the principal idea of nonparametric kernel estimation of $m(x)$, i.e. to construct a weighted average of those observations (X_i, Y_i) of which X_i happens to fall into an asymptotically shrinking neighborhood of x. The shrinkage of such a neighborhood is usually parameterized by a sequence of *bandwidths* h_n tending to zero, whereas the shape of the neighborhoods is given by a real *kernel* function K.

Motivated by classical procedures we define the following sequential estimator of a zero of m,

$$(1) \qquad Z_{n+1} = Z_n - a_n h_n^{-1} K((Z_n - X_n)/h_n) Y_n, \quad n \geq 1.$$

Here Z_1 denotes an arbitrary starting random variable with finite second moment and $\{a_n\}$ is a sequence of positive constants tending to zero. In fact, the sequence $\{Z_n\}$ will converge under our conditions to the (unique) zero of

$$\tilde{m}(x) = \int y f(x,y) dy = m(x) f_X(x),$$

where $f_X(x)$ denotes the marginal density of X, but an assumption about f_X ensures that the zero of the two functions m and $\tilde{m}$ is identical.

Under mild conditions we show consistency (almost surely and in quadratic mean) and asymptotic normality of $\{Z_n\}$. An asymptotic bias term (depending on the smoothness of m) shows up, if the bandwidth sequence tends to zero at a specific rate. Fixed width confidence intervals are constructed, using a suitable stopping rule based on estimates of the variance of the asymptotic normal distribution.

Our arguments can be extended to the problem of estimating extremal values of the regression function m. Note that $m = \tilde{m}/f_X$ and therefore $m' = \tilde{r}/f_X^2$, where

$$\tilde{r}(x) = f_X(x) \int y \frac{\partial}{\partial x} f(x,y) dy - \tilde{m}(x) \frac{\partial}{\partial x} f_X(x).$$

Under a suitable assumption the problem of finding an extremum of m is equivalent to finding a (unique) zero of the function r. So it is reasonable to apply a procedure similar to (1). Additional difficulties turn up since f_X has to be estimated separately. We propose to perform the estimation by an additional i.i.d. sequence $\{\bar{X}_i\}$ with the same distribution as X. Define

1

$$(2) \qquad Z'_{n+1} = Z'_n - a_n h_n^{-3} K((Z'_n - \bar{X}_n)/h_n) K'((Z'_n - X_n)/h_n) Y_n$$

$$+ a_n h_n^{-3} K'((Z'_n - X_n)/h_n) K((Z'_n - X_n)/h_n) Y_n, \quad n \geq 1$$

We shall prove that $\{Z'_n\}$ is consistent and asymptotically normally distributed. Fixed width confidence intervals are computed by the same technique as for $\{Z_n\}$.

If we knew f_X the algorithm (2) would simplify, the additional $\{X_i\}$ are obsolete in this case, here we propose

$$(3) \qquad Z'_{n+1} = Z'_n - a_n h_n^{-2} K'((Z'_n - X_n)/h_n) Y_n f_X(Z'_n)$$

$$+ a_n h_n^{-1} K((Z'_n - X_n)/h_n) Y_n f'_X(Z'_n), \quad n \geq 1.$$

The additional difficulty of estimating simultaneously f_X didn't occur in the case of estimating zeros, since the problem for m could be transferred to the equivalent problem for $\tilde{m}$, which does not involve f_X. In practice the additional i.i.d. sequence $\{X_i\}$ could be constructed by sampling in pairs and discarding the Y observations of one element. This results in some loss of efficiency but makes the practical application possible with the data at hand. Another proposal that we would like to make is related to the boot-strap. From the first N observations, a density estimate $\hat{f}_X$ of f_X could be constructed and then the algorithm (2) could be started with $\{X_i\}$ distributed with density $\hat{f}_X$. A third possibility is to plug in $\hat{f}_X$ into the algorithm (3). We did not investigate the last mentioned procedures.

An alternative way of defining an estimator of the zero of the regression function m could be to construct an estimate of the whole function and then to empirically determine an observed zero as an estimate. This procedure would be time consuming in the case of sequential observation of the date, since for every new observation the whole function has to be constructed whereas our procedure just keeps one number in memory and updates that number due to the formal presription (1). Also in cases where an enormous amount of data has to be processed, an estimate of a zero based on the estimate of the whole regression function seems to be inadvisable since all the data has to be stored in the memory at a time.

Related work was done by Revesz (1977) and Rutkowski (1981, 1982) who applied stochastic approximation methods to the estimation of m at a fixed point. Our derivation of fixed width confidence intervals was inspired by the papers of Chow and Robbins (1965), McLeish (1976) and Stute (1983). The author last mentioned used in the field of density estimation the kernel estimation technique that introduces a localizing effect which makes classical methods, such as Venter's (1966), applicable.

The rest of the paper is organized as follows. Section 2 contains the results and gives the consistency proof for $\{Z_n\}$. In section 3 we present the results of some simulations and an application of of $\{Z_n\}$ to some real data. In the last section we give the rest of the proofs.

2. Results

A crucial assumption that makes the problem identifiable through $\bar{m}$ [resp. r] is the following.

(A1)   The marginal density f_X of X is positive.

The speed of convergence of $\{a_n\}$ and $\{h_n\}$ is controlled by

$$(A2.1) \qquad \sum_{n=1}^{\infty} a_n = \infty \quad , \quad \sum_{n=1}^{\infty} a_n h_n < \infty$$

$$(A2.2) \qquad \sum_{n=1}^{\infty} a_n^2 h_n^{-2} < \infty$$

$$(A2.3) \qquad \sum_{n=1}^{\infty} a_n^2 h_n^{-4} < \infty$$

The zero Θ_0 of $m(x)$ (and of $\bar{m}(x)$) is identified by

$$(A3) \qquad \inf_{|x-\Theta_0| \geq \epsilon} (x - \Theta_0)\bar{m}(x) > 0 \quad \text{for all} \quad \epsilon > 0.$$

Smoothness of $\bar{m}$ is guaranteed by

(A4.1) $\bar{m}$ is Lipschitz continuous;

(A4.2) $\bar{m}$ is differentiable in a neighborhood of Θ_0 such that

$$\bar{m}'(\Theta_0) > 1/4;$$

(A4.3) $\bar{m}$ is twice continuously differentiable.

The kernel function K has to satisfy the following conditions.

(A5.1) K is bounded and

$$\int K(u)du = 1, \int uK(u)du = 0, \int u^2 K(u)du < \infty$$

(A5.2) K is differentiable and

$$\lim_{|u| \to \infty} |uK(u)| = 0, \int |u| K'^2(u)du < \infty$$

3

(A5.3) K is twice differentiable and

$$\lim_{|u|\to\infty} |uK'(u)| = 0, \int |u| K''^2(u)du < \infty.$$

The joint density $f(x,y)$ has to be smooth in its first argument.

(A6.1) $|f(x,y) - f(z,y)| \le |x - z| g_1(y)$ such that $\int (y^2 + 1)g_1(y)dy < \infty$.

(A6.2) $\frac{\partial^2}{\partial x^2} f(x,y)$ is continuous and

$\left|\frac{\partial}{\partial x}f(u,y) - \frac{\partial}{\partial x}f(v,y)\right| \le |u - v| g_2(y)$

such that $\int (|y| + 1)g_2(y)dy < \infty$.

Moment assumptions are

(A7) $EY^4 < \infty$ and $\sup_x E(Y^2|X = x)$.

We have split up the assumptions into several subparts since we will use the subparts separately. The consistency of $\{Z_n\}$ is shown in

<u>Theorem 1.</u>   <u>Assume $(A1), (A2.1), (A2.2), (A3), (A4.1), (A5.1), (A7)$. Then $\{Z_n\}$ converges to Θ_0 almost surely and in the quadratic mean.</u>

Since the proof of this theorem is very simple and exemplifies the combination of the kernel method together with stochastic approximation arguments we would like to give it here. The proofs of the following results are delayed to section 4.

Write

$$Z_{n+1} = Z_n - a_n \tilde{m}(Z_n) + a_n V_n$$
$$V_n = \tilde{m}(Z_n) + K_h(Z_n - X_n)Y_n$$
where $K_h(u) = h_n^{-1} K(u/h_n)$.

Let $\mathcal{B}_n = \sigma\{Z_1, Z_2, \ldots, Z_n\}$. Condition (A4.1) implies that

$$E(V_n|\mathcal{B}_n) = O(h_n)   a.s.$$
$$E(V_n^2) = O(E(Z_n - \Theta_0)^2) + O(h_n^{-2}).$$

Observe that with (A3) and a Lipschitz constant $L_{\tilde{m}}$

$$
\begin{aligned}
(Z_{n+1} - \Theta_0)^2 &= (Z_n - \Theta_0)^2 - 2a_n \tilde{m}^2(Z_n)(Z_n - \Theta_0) \\
&\quad + a_n^2 \tilde{m}^2(Z_n) + 2a_n V_n(Z_n - \Theta_0 - a_n \tilde{m}(Z_n)) \\
&\quad + a_n^2 V_n^2 \\
&\le (1 + a_n^2 L_{\tilde{m}}^2)(Z_n - \Theta_0)^2 + a_n^2 V_n^2 \\
&\quad + 2a_n V_n(Z_n - \Theta_0 - a_n \tilde{m}(Z_n)).
\end{aligned}
$$

4

Hence by (A7),

$$E((Z_{n+1} - \Theta_0)^2 | \mathcal{B}_n) \leq (1 + a_n^2 L_{\bar{m}}^2)(Z_n - \Theta_0)^2$$
$$+ O(h_n)a_n(1 + a_n L_{\bar{m}})|Z_n - \Theta_0|$$
$$+ a_n^2 E(V_n^2 | \mathcal{B}_n)$$
$$\leq (1 + \beta_n)(Z_n - \Theta_0)^2 + \delta_n.$$

where

$$\beta_n = O(h_n^{-2} a_n^2 + h_n a_n + a_n^2)$$
$$\delta_n = O(h_n a_n + h_n^{-2} a_n^2),$$

if we use the simple inequalities

$$|Z_n - \Theta_0| \leq 1 + |Z_n - \Theta_0|^2$$
$$E(V_n^2 | \mathcal{B}_n) \leq 2\bar{m}^2(Z_n) + O(h^{-2}) \sup_x E(Y^2 | X = x)$$

Note that by (A2.1), (A2.2) $\quad \sum \beta_n; \sum \delta_n < \infty.$

The assertion follows now from Venter (1966), Theorem 1. Nixdorf (1982), Theorem 1.1.2 has given a corrected version.

The asymptotic normality is shown in

Theorem 2.   Assume $(A1), (A3), (A4.2), (A4.3), (A5.1)$.
Let $a_n = n^{-1}, h_n = n^{-\gamma}, 1/5 \leq \gamma \leq 1/2$.
Then

$$n^{\frac{1-\gamma}{2}} \{Z_n - \Theta_0\} \to N(b(\gamma), \sigma^2(\gamma))$$

where

$$b(\gamma) = 0 \qquad \text{if} \qquad 1/5 < \gamma < 1/2$$
$$= \bar{m}''(\Theta_0) \int u^2 K(u) du / (2\bar{m}'(\Theta_0) - 1 + \gamma) \qquad \text{if} \quad \gamma = 1/5;$$
$$\sigma^2(\gamma) = \int K^2 \int y^2 f(\Theta_0, y) dy / (2\bar{m}'(\Theta_0) - 1 + \gamma)$$

Fixed width asymptotic confidence intervals for the unknown parameter Θ_0 are constructed via estimators of $b(\gamma)$ and $\sigma^2(\gamma)$.

Estimators of $\int y^2 f(\Theta_0, y) dy, \bar{m}'(\theta_0), \bar{m}''(\Theta_0)$ are

$$(4) \qquad S_{1n} = n^{-1} \sum_{i=1}^{n} K_{h_i}(Z_i - X_i) Y_i^2,$$

$$(5) \qquad S_{2n} = n^{-1} \sum_{i=1}^{n} K'_{h_i}(Z_i - X_i) Y_i,$$

$$S_{3n} = n^{-1} \sum_{i+1}^{n} K''_{h_i}(Z_i - X_i) Y_i,$$

respectively.

An estimator for the asymptotic variance $\sigma^2(\gamma)$ is therefore

$$s_n = \int K^2 S_{1n}/(2S_{2n} - 1 + \gamma) \quad \text{if} \quad 2S_{2n} - 1 + \gamma > 0,$$
$$= 0 \qquad\qquad\qquad\qquad \text{else}$$

So the following stopping rule seems reasonable.

$$(6) \qquad N(d) = inf\{n \in \mathbb{N} \mid s_n + n^{-1} \le n^{1-\gamma} d^2/z_{\alpha/2}\}.$$

where $z_{\alpha/2}$ is the $(1 - \alpha/2)-$ quantile of the standard normal distribution.

The fixed width confidence intervals are constructed via

<u>Theorem 3.</u> <u>Let</u> $a_n = n^{-1}, h_n = n^{-\gamma}, 1/5 \le \gamma \le 1/3$ <u>and assume</u> $(A1), (A3), (A4.2), (A4.3),$ $(A5.1), (A5.2)$. <u>Then if</u> $N(d)$ <u>is defined as in</u> (6) <u>for some</u> $0 < \alpha < 1$, <u>as</u> $d \to 0$

$$N(d)^{\frac{1-\gamma}{2}} \{Z_{N(d)} - \Theta_0\} \to^{\mathcal{L}} N(b(\gamma), \sigma^2(\gamma)).$$

In the case $1/5 < \gamma < 1/3$ an asymptotic confidence interval of fixed length $2d$ and asymptotic coverage probability $1 - \alpha$ is given by

$$[Z_{N(d)} - d, Z_{N(d)} + d].$$

For $\gamma = 1/5$ the bias can be estimated by

$$b_n = \int u^2 K(u) du\, S_{3n}/(2S_{2n} - 1 + \gamma).$$

Then with $H_n = Z_n - n^{\frac{-1+\gamma}{2}} b_n$ an asymptotic confidence interval is given by $[H_{N(d)} - d, H_{N(d)} + d]$.

6

<u>Remark 1.</u>   The range of γ had to be reduced to $1/5 \leq \gamma < 1/3$ since otherwise S_{2n} would no longer be a consistent estimator of $\tilde{m}'(\Theta_0)$.

<u>Remark 2.</u>   It will be seen in the proof of Theorem 3 that, as $d \to 0, N(d)/b(d) \to 1$ almost surely where $b(d) = inf\{n \in \mathbb{N} \mid \sigma^2(\gamma) \leq n^{1-\gamma}d^2/z_{\alpha/2}^2\}$. Therefore $N(d)$ exhibits the following limit behavior,

$$d^{2/(1-\gamma)}N(d) \to (\sigma^2(\gamma))^{1/(1-\gamma)}z_{\alpha/2}^{2/(1-\gamma)},$$

as $d \to 0$.

The analysis of the sequential procedure $\{Z_n'\}$ is quite analogeous to that of $\{Z_n\}$. we define the (unique) zero of $\tilde{r}$ as Θ_M.

<u>Theorem 4.</u>   <u>Assume</u> $(A1),(A2.1),(A2.3),(A5.1),(A5.3),(A6.1),(A7)$ <u>and let</u> $(A3),(A4.1)$ <u>be fulfilled with $\tilde{r}$ in the place of $\tilde{m}$. Then $\{Z_n'\}$ converges to Θ_M almost surely and in the quadratic mean.</u>

The next theorem gives the asymptotic normality of $\{Z_n'\}$.

<u>Theorem 5.</u>   <u>Let</u> $a_n = n^{-1}$ <u>and</u> $h_n = n^{-\gamma}, 1/6 < \gamma < 1/5$ <u>then under</u> $(A1),(A5.1),(A5.3),$ $(A6.2),(A7)$ <u>and</u> $(A3),(A4.2),$ <u>with $\tilde{r}$ in the place of $\tilde{m}$.</u>
<u>Then</u>

$$n^{\frac{1-4\gamma}{2}}\{Z_n' - \Theta_M\} \to^{\mathcal{L}} N(0,\sigma_M^2(\gamma)),$$

<u>where</u>

$$\sigma_M^2(\gamma) = f_X(\Theta_M)\int y^2 f(\Theta_M,y)dy \int K^2 \int (K')^2/(2\tilde{r}'(\Theta_M) - 1 + 4\gamma)$$

<u>Remark 3.</u>   For simplicity of presentation we didn't arrange for a wider range of γ such that an asymptotic bias term occurs. If $\tilde{r}$ is twice continuously differentiable then the range of allowable exponents can be extended to $1/8 \leq \gamma < 1/4$. The discussion would be in analogy to Theorem 2 with $\tilde{r}$ in the place of $\tilde{m}$.

Estimators for the numerator and denominator of $\sigma_M^2(\gamma)$ are constructed in the following way.

$$S_{1n}' = n^{-1}\sum_{i=1}^{n} K_{h_i}(Z_i - \tilde{X}_i)n^{-1}\sum_{i=1}^{n} K_{h_i}(Z_i - X_i)Y_i^2$$

is an estimator for $f_X(\Theta_M)\int y^2 f(\Theta_m,y)dy$, whereas

$$S_{2n}' = n^{-1}\sum_{i=1}^{n} K_{h_i}(Z_i - X_i)n^{-1}\sum_{i=1}^{n} K_{h_i}''(Z_i - x_i)Y_i$$
$$-n^{-1}\sum_{i=1}^{n} K_{h_i}(Z_i - X_i)Y_i n^{-1}\sum_{i=1}^{n} K_{h_i}''(Z_i - \tilde{X}_i)$$

7

converges under our assumptions to $\tilde{r}'/\Theta_M$, almost surely. Define

$$s_{n,M} = \int K^2 \int (K')^2 S'_{1n}/(2S'_{2n} - 1 + 4\gamma)$$

$$N'(d) = \inf\{n \in \mathbb{N} \mid s_{n,M} + n^{-1} \leq n^{1-2\gamma}d^2 Z_{\alpha/2}\}.$$

Then parallel to Theorem 3 we have

<u>Theorem 6.</u> Let $a_n = n^{-1}$ and $h_n = n^{-\gamma}, 1/6 \leq \gamma < 1/5,$ <u>and let the conditions of Theorem 5</u> be fulfilled. Then, as $d \to 0$

$$N'(d)^{1/(2-2\gamma)}\{Z'_{N'(d)} - \Theta_M\} \to^{\mathcal{L}} N(0, \sigma_M^2(\gamma)).$$

3. Monte Carlo Study and an Application

In this section we report the results of a Monte Carlo experiment comparing the performance of our sequential procedure when some of the involved parameters are tuned at different levels. We also report an application of the algorithm (1) to some real data.

The basic experiment to assess the accuracy of Theorem 3 consisted of 200 Monte Carlo replications with the numbers $N(d), Z_{N(d)}$ and $S_{N(d)}$ to be reported. The joint probability density function $f(x,y)$ that we used was $f(x,y) = I_{[0,1]}(x)\sigma_*^{-1}\varphi((y - m(x))/\sigma_*)$, φ the probability density function of a standard normal distribution and $m(x) = -a\{(1-x)^2 - 1/4\}$ for $a = 4, 8$ was the regresssion curve. We report the result for $Z_1 = 0.45$ (Table 1) and for $Z_1 = 0.2$ (Table 2). The parameter α was set to $\alpha = 0.05$. The zero that was to be estimated was $\Theta_0 = 1/2$ and two different values of d and σ_* were fixed, namely $d = 0.05, 0.1$ and $\sigma_* = 0.1, 1.0$. As the kernel K we have chosen the Epanechnikov kernel $K(u) = 3/4(1 - u^2)$ for $|u| \leq 1$ and $K(u) = 0$ for $|u| > 1$. the sequence of bandwidths was set to $h = h_n = n^{-\gamma}$, $\gamma = 0.21$. In Table 1 the results for the starting point $Z_1 = 0.45$ are shown. The figures of Table 1 indicate that the fixed accuracy result

Table 1 about here

given in Theorem 3 yields a good approximation of Θ_0 even for $d = 0.1$. This is seen from the counts in the $Z_{N(d)}$- column. It is indicated there how many times (from 200 Monte Carlo trials) the true parameter $\Theta_0 = 1/2$ was in the confidence interval $[Z_{N(d)} - d, Z_{N(d)} + d]$. As a measure of spread we added the quantiles Q_{95} and Q_5 in the third and fourth column of each entry. A small paradox occurs when we compare the figures for different values of a. It is expected that the procedure (1) stops earlier with $a = 8$ than with $a = 4$, since the higher derivative in the zero should speed up the convergence of $\{Z_n\}$ to Θ_0. In both Table 1 and Table 2 it is seen that the average of the stopping times

Table 2 about here

over 200 Monte Carlo runs) is considerably higher for $a = 8$ and $\sigma_* = 0.1$ than for $a = 4$ and $\sigma_* = 0.1$. This effect is due to the crude approximation $var(Y|X = x) \approx \sigma^2$, $x \approx \Theta_0$, as can be seen from the figures for $S_{N(d)}$. In the case of $a = 8$ the statistic $S_{N(d)}$ considerably overestimates the true asymptotic variance $\sigma(\gamma)$. For comparison we list some correct $\sigma(\gamma) = \sigma(\sigma_*, a, \gamma)$. For instance, $\sigma(0.1, 4, 0.21) = 0.00083$ whereas $\sigma(0.1, 8, 0.21) = 0.00039$.

In a small application we took the sequence of random variables $\{(X_i, Y_i)\}$, X_i =age, Y_i =weight

9

of female corpses) which was gathered from 1969 to 1981 by the Institute of Forensic Medicine of Heidelberg. It is an interesting question in forensic medicine to estimate the mean age from the weight of unknown corpses. We restricted our attention to the ages between 0 and 20 years in order to fulfill assumption $(A3)$. We put $m_0 = 40$ kg, and we applied the procedure (1) and ended with different starting values Z_1 at $Z_{N(d)} = 11.6$ years and $N(d) = 563$, for $d = 0.1$ and $N(d) = 224$ for $d = 0.2 (Z_1 = 0.4)$. A plot of the first 732 data pairs, restricted to ages between 0 and 20 years, should illustrate the accuracy of $Z_{N(d)}$ (Figure 1).

4. Proofs

The theorems are proved by a functional central limit theorem given by Berger (1980), who extended a result of Walk (1977), that made it applicable in our setting. Lemma 1 describes the asymptotic behavior of

$$(7) \qquad W_n(t) = n^{-1/2} R_{\lfloor nt \rfloor} + n^{-1/2}(nt - \lfloor nt \rfloor)\{R_{\lfloor nt \rfloor + 1} - R_{\lfloor nt \rfloor}\}, \quad 0 \le t \le 1,$$

where

$$R_k = k^{1/2}[k^{\frac{1-\gamma}{2}}(Z_k - \Theta_0) - b(\gamma)], \quad k \in \mathbb{N}.$$

<u>Lemma 1</u>. Let the conditions of Theorem 3 be satisfied, then $W_n(t)$, as defined in (7) converges weakly in $C[0, 1]$ to the Gaussian process

$$(8) \qquad G_1(t) = \int K^2 \int y^2 f(\Theta_0, y) dy \int_{(0,1]} u^{\tilde{m}'(\Theta_0) - (2+\gamma)/2} dW(tu), \quad 0 \le t \le 1,$$

where W is the standard Wiener process starting at 0.

<u>Proof.</u> Define

$$\mathcal{B}_n = \sigma\{Z_1, \ldots, Z_n\}$$

$$Z_{n+1} - \Theta_0 = (1 - B_n/n)(Z_n - \Theta_0) + n^{\frac{-(2-\gamma)}{2}} \tilde{V}_n + n^{\frac{-(2-\gamma)}{2}} T_n$$

where

$$\tilde{V}_n = h^{-1/2} E\{K(\frac{Z_n - X_n}{h})Y_n | \mathcal{B}_n\} - h^{-1/2} K(\frac{Z_n - X_n}{h})Y_n,$$

$$T_n = n^{\frac{1-\gamma}{2}}\{\tilde{m}(Z_n) - E[K_h(Z_n - X_n)Y_n | \mathcal{B}_n]\}$$

and $\{B_n\}$ is a sequence of random variables converging almost surely to $\tilde{m}'(\Theta_0)$ such that $B_n(Z_n - \Theta_0) = \tilde{m}(Z_n)$. Such a sequence exists because $\tilde{m}$ is differentiable in Θ_0 and $Z_n \to \Theta_0$ almost surely by Theorem 1. The assumption on a_n and h_n imply that $T_n \to 1/2 \int u^2 K(u) du \tilde{m}''(\Theta_0)$. Note that $E(\tilde{V}_n | \mathcal{B}_n) = 0$ and that by $(A7)$ and $(A6.1)$,

$$E(\tilde{V}_n^2 | \mathcal{B}_n) \to \int K^2 \int y^2 f(\Theta_0, y) dy, \text{ almost surely },$$

$$E(\tilde{V}_n^2) = O(1).$$

Furthermore we have for all $r > 0$

$$E(\tilde{V}_n^2 I(\tilde{V}_n^2 \ge rn | \mathcal{B}_n) \le O(h^{-2}) P(\tilde{V}_n^2 \ge rn | \mathcal{B}_n)$$

$$\le O(h^{-2}n^{-1}) = o(1) \quad \text{almost surely.}$$

11

The lemma follows now from the generalization of a theorem of Walk (1977) , given by Berger (1980).

□

The following lemma gives an analogeous result for the Kiefer Wolfowitz type sequence $\{Z'_n\}$ defined in (2)

Lemma 2.   Let the conditions of Theorem 6 be satisfied. Define $W_n(t)$ as in lemma 1 but with $\tilde{r}$ in the place of $\tilde{m}$ and

$$R_k = k^{1/2} k^{\frac{(1-\gamma)}{2}} (Z_k - \Theta_M).$$

Then $W_n(t)$ converges weakly in $C[0,1]$ to the Gaussian process

$$G_2(t) = f_X(\Theta_M) \int y^2 f(\Theta_M, y) dy \int K^2 \int (K')^2 \int_{(0,1]} u^{\tilde{r}'(\Theta_M)-(1-\gamma/2)} dW(tu), \quad 0 \leq t \leq 1.$$

Proof of Theorem 2.   Use Lemma 1 and evaluate $G_1(t)$ at $t = 1$.

Prooof of Theorem 3.   Define the sequence

$$b(d) = inf\{n \in \mathbb{N} \mid \sigma^2(\gamma) \leq n^{1-\gamma} d^2 / z^2_{a/2}\}.$$

The estimators S_{1n}, S_{2n} defined in (4), (5) converge to $\int y^2 f(\Theta_0, y) dy, \tilde{m}'(\Theta_0)$ respectively. This entails that, as $d \to 0$,

$$N(d)/b(d) \to 1 \text{ almost surely.}$$

Now apply Lemma 1.

Proof of Theorem 4.   Like the proof of Theorem 1.

Proof of Theorem 5.   Use Lemma 2 and evaluate $G_2(t)$ at $t = 1$.

Proof of Theorem 6.   Similar to the proof of Theorem 3.

12

References

[1] Berger [1980]: *A note on the invariance principle for stochastic approximation in a Hilbert space*. Manuscript.

[2] Chow, Y.S. and Robbins, H. [1965]: *On the asymptotic theory of fixed-width sequential confidence intervals for the mean*. Ann. Math. Statist. 36, 457 462.

[3] McLeish, R. [1976]: *Functional and random central limit theorems for the Robbins-Monro process*. J. Appl. Prob. 13, 148-154.

[4] Nixdorf, R. [1982]: *Stochastische Approximation in Hilberträumen durch endlichdimensionale Verfahren*. Mitt. Math. Sem. Gießen, Heft 154.

[5] Revesz, P. [1977]: *How to apply the method of stochastic approximation in the non-parametric estimation of a regression function*. Math. Oper. Statist. 8, 119-126.

[6] Robbins, H. and Monro, S. [1951]: *A stochastic approximation method*. Ann. Math. Statist. 22, 400-407.

[7] Rutkowski, L. [1981]: *Sequential estimates of a regression function by orthogonal series with applications in discrimination*. Lecture Notes in Statistics 8, Springer Verlag, 236-244.

[8] Rutkowski, L. [1982]: *On-line identification of time- varying systems by nonparametric techniques*. IEEE trans. Autom Control. 27, 228-230.

[9] Stute, W. [1983]: *Sequential fixed-with confidence intervals for a non-parametric density function*. Z. Wahrsch. 62, 113-123.

[10] Venter, J.H. [1966]: *On Dvoretzky stochastic approximation theorems*. Ann. Math. Statist. 37, 1534-1544.

[11] Walk, H. [1977]: *An invariance principle for the Robbins- Monro process in a Hilbert space*. Z. Wahrsch. 39, 135-150.

13

Table 1

ζ₀ = .45, γ = .21, α = .05

	d	N(d)				Z_N(d)					D_N(d)				
		mean	std	Q5	Q95	mean	std	Q5	Q95	counts	mean	std	Q5	Q95	m'(θ)
.1	.05	137	10	120	156	.518	.021	.483	.553	188	.024	.002	.02	.028	4
.1	.05	229	17	200	259	.515	.019	.483	.547	194	.043	.003	.037	.048	8
.1	.1	43	3.5	38	50	.519	.03	.469	.574	199	.027	.005	.018	.035	4
.1	.1	62	6.7	51	74	.517	.039	.449	.582	196	.05	.007	.038	.062	8
.1	.05	642	97	496	806	.51	.025	.467	.548	188	.105	.013	.127	.084	4
.1	.05	469	46	394	551	.515	.023	.475	.559	181	.081	.006	.07	.093	8
.1	.1	129	37	76	207	.52	.054	.429	.61	184	.11	.028	.065	.168	4
.1	.1	103	18	72	133	.525	.043	.461	.596	192	.09	.015	.062	.114	8

Table 2

ζ₀ = .2, γ = .21, α = .05

	d	N(d)				Z_N(d)					D_N(d)				
		mean	std	Q5	Q95	mean	std	Q5	Q95	counts	mean	std	Q5	Q95	m'(θ)
.1	.05	163	12	141	183	.517	.018	.485	.547	192	.03	.002	.025	.034	4
.1	.05	255	17	227	283	.518	.019	.484	.552	189	.04	.003	.042	.052	8
.1	.1	56	7	46	70	.513	.027	.464	.561	199	.04	.007	.031	.059	4
.1	.1	74	8	61	90	.515	.036	.456	.593	196	.064	.007	.05	.077	8
.05	.05	646	83	518	uO2	.516	.026	.471	.562	174	.105	.011	.088	.126	4
.05	.05	479	44	417	550	.512	.023	.475	.55	188	.082	.006	.073	.093	8
.1	.1	139	31	99	192	.515	.044	.437	.595	192	.118	.024	.077	.158	4
.1	.1	118	19	86	146	.522	.042	.454	.59	193	.102	.015	.074	.125	8
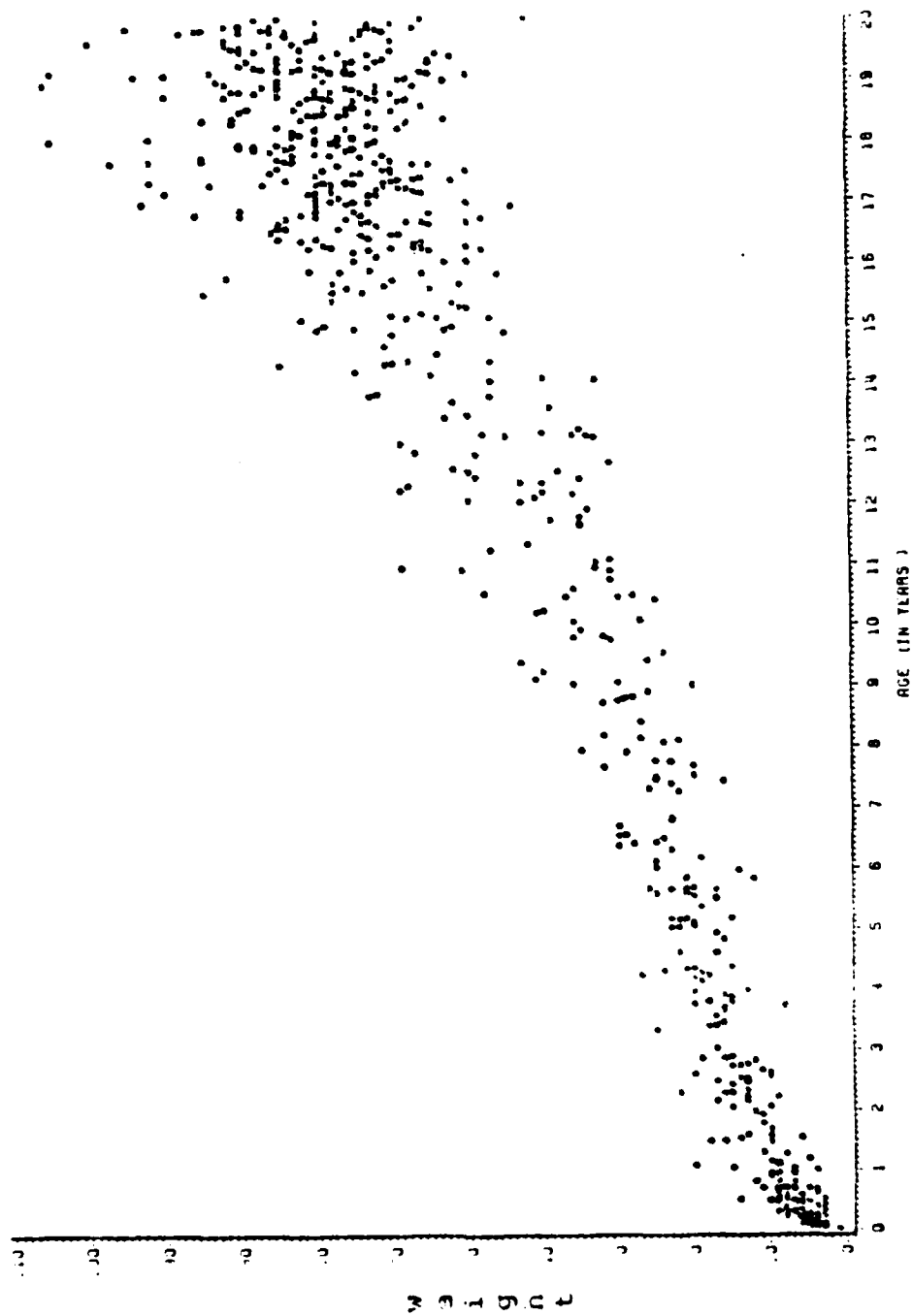
Weight and age for 732 female corpses

Figure 1.

END

DTIC

8-86